

Gencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:36 ; Search time 170.72 Seconds

(without alignments)
17.102 Million cell updates/sec

Title: US-09-331-631a-5_COPY_33_75
Perfect score: 248
Sequence: 1 NQEDPQTECQQCQRRCRQE.....RQQYCQRRCKEICEEEB 43
Scoring table: BLOSUM62
Gapop 10.0 , **Gapext** 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_66;*

1: pir1;*
2: pir2;*
3: pir3;*
4: pir4;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	110	44.4	566	2 S22477	vicilin precursor
2	105	42.3	588	1 FWCNAB	alpha-globulin B p
3	104	41.9	509	2 S03059	alpha-globulin typ
4	96	38.7	411	2 T29475	hypothetical prote
5	94	37.9	605	2 S03398	alpha-globulin typ
6	78.5	31.7	317	2 T28592	hypothetical prote
7	78.5	31.7	600	2 T18593	hypothetical prote
8	77	31.0	572	2 T28880	hypothetical prote
9	77	31.0	1513	2 T23681	hypothetical prote
10	75.5	30.4	242	2 T26999	hypothetical prote
11	75	30.2	388	2 T31087	hypothetical prote
12	75	30.2	388	2 T31888	hypothetical prote
13	74	29.8	438	2 T31899	hypothetical prote
14	74	29.8	445	2 T31898	hypothetical prote
15	73.5	29.4	330	2 T25169	hypothetical prote
16	73	29.4	330	2 T31559	hypothetical prote
17	72.5	29.2	335	2 T31560	hypothetical prote
18	72.5	28.8	335	2 T31561	hypothetical prote
19	71.5	28.8	335	2 JC5557	hypothetical prote
20	71	28.6	47	2 JC5557	arginine/glutamate
21	68.5	28.6	356	2 S31574	hypothetical prote
22	68	27.4	154	2 T27967	hypothetical prote
23	67.5	27.2	314	2 T21686	hypothetical prote
24	66.5	26.8	273	2 T20930	hypothetical prote
25	66.5	26.8	637	2 S35221	globulin Bg1 prec
26	66.5	26.8	654	2 T30136	hypothetical prote
27	66	26.6	111	2 T20452	hypothetical prote
28	65.5	26.4	33	2 A1822	antimicrobial pept
29	65.2				hypothetical prote

ALIGNMENTS

RESULT	1
S22477	vicilin precursor - cacao

C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050

R:McHenry, L.; Fritz, P.J.
Plant Mol. Biol., Biol. 18, 1173-1176, 1992

A:Title: Comparison of the structure and nucleotide sequence of vicilin genes of coca
A:Reference number: S22477; MUID:92288309

A:Molecule type: mRNA
A:Accession: S22477

A:Molecule type: DNA
A:Residues: 1-452 <MC2>

A:Cross-references: EMBL:X62625
A:Cross-references: Fasta
A:Cross-references: Genbank
A:Cross-references: EMBL:X62625

A:Keywords: seed; storage protein
C:Superfamily: glycinin
F:1-24-Domain: signal sequence #status predicted <SIG>
F:25-566/Product: vicilin #status predicted <MAT>

Query	Match	Best Local Similarity	Score	Length
Matches	19; Conservative	47.5%	110	566;
QY	2 QBDPOTECQQCQRRCQESPRODQYCORCRKEICEEE 41			
Db	78 REELQRYQQGRCBQQQREQQCQRCKWEQYKEQE 117			

RESULT 2

FWCNAB
alpha-globulin B precursor (clone C72) - upland cotton
N:Alternate names: seed storage protein; vicilin precursor

C:Species: Gossypium hirsutum (upland cotton)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: A30838; S06911

R:Chian, C.A.; Pyle, J.B.; Leacock, A.B.; Dure III, L.
Plant Mol. Biol., 7, 475-489, 1986

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVII
A:Reference number: A30838
A:Accession: A30838

A:Molecule type: mRNA
A:Residues: 1-588 <CHI>

A:Cross-references: GB:MI6891; MUID:9167374; PIDN:AAA33071.1; PID:9167375
A:Experimental source: var. Coler 201
R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

Plant Mol. Biol. 9, 533-546, 1987
 A; Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
 A; Reference number: S06398
 A; Accession: S06391
 A; Status: not compared with conceptual translation
 A; Molecule type: DNA
 A; Residues: 1-81 <CH2>
 C; Comment: This is a seed storage protein.
 C; Superfamily: glycaminin
 C; Keywords: glycoprotein; seed; storage protein
 F; 1-25/Domain: signal sequence #status predicted <SIG>
 F; 26-588/Product: alpha-globulin storage protein #status predicted <MAT>
 F; 417/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F; 417/Binding site: carbohydrate (Asn) (covalent) #status predicted
 QY 2 QEDPOTECQQ-CQRKRCRQQESDPROQQYCQRCRKEICEEEFY 43
 Db 138 QQPRQQCQDCQANCPQQ--QQQQQQQQCQTCQSBDQY 177
 RESULT 3
 S08059 alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
 N; Alternate names: seed storage protein
 C; Species: *Gossypium hirsutum* (upland cotton)
 C; Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
 C; Accession: S08059
 R; Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
 R; Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
 Plant Mol. Biol. 9, 533-546, 1987
 A; Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
 A; Reference number: S06398
 A; Accession: S06398
 A; Status: not compared with conceptual translation
 A; Molecule type: DNA
 A; Residues: 1-605 <CHL>
 C; Superfamily: glycaminin
 F; 1-24/Domain: signal sequence #status predicted <SIG>
 F; 25-605/Product: alpha-globulin type A #status predicted <MAT>
 QY 2 QEDPOTECQQCQRQQESDPROQQYCQRCRKEICEEE 40
 Db 115 QQQPDKQFKECQQRQWQEPRKRQVKECREQYED 153
 RESULT 4
 T18592 hypothetical protein AC3;3 - *Caenorhabditis elegans*
 C; Species: *Caenorhabditis elegans*
 C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C; Accession: T18592
 R; McMurray, A. submitted to the EMBL Data Library, April 1996
 A; Reference number: Z18955
 A; Accession: T18592
 A; Status: preliminary; translated from GB/EMBL/DDBJ
 A; Molecule type: DNA
 A; Residues: 1-425 <WIL>
 A; Cross-references: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3
 A; Experimental source: clone AC3
 C; Genetics:
 A; Gene: CESP:AC3.3
 A; Map position: 5
 A; Introns: 18/3
 C; Superfamily: gliadin
 A; Reference number: 220623
 A; Accession: T29475
 A; Status: preliminary; translated from GB/EMBL/DDBJ
 A; Molecule type: DNA
 A; Residues: 1-411 <BRA>
 A; Cross-references: EMBL:UB0455; PIDN:ABA37887.1; GSPDB:GN00020; CESP:T01D1.6
 A; Experimental source: strain Bristol N2; clone T01D1
 C; Genetics:
 A; Gene: CESP:T01D1.6
 A; Map position: 2
 A; Introns: 25/3; 304/3
 C; Superfamily: gliadin
 QY 2 QEDPOTECQQCQRQQESDPROQQYCQRCRKEICEEEFY 38
 Db 255 QQPSPAPQCQQCQNTCQQAAPVCQQCAPCQQQSAAPACQQCQTSCQTCQGQCGCTPOCQ 314
 QY 39 E 39
 Db 315 O 315

RESULT 7
T18593 hypothetical protein AC3.4 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999
C;Accession: T18593
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z18995
A;Accession: T18593
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-600 <WIL>
A;Cross-references: EMBL:Z71177; PIDN:CAA94868.1; GSPDB:GN00023; CESP:AC3.4
A;Experimental source: clone AC3
C;Genetics:
A;Gene: CESP:AC3.4
A;Map position: 5
A;Introns: 292/2; 374/3; 572/2

Query Match 31.7%; Score 78.5; DB 2; Length 600;
Best Local Similarity 27.9%; Pred. No. 0.47; Mismatches 9; Indels 23; Gaps 2;
Matches 17; Conservative 9; Mismatches 12; Indels 23; Gaps 2;

QY 2 QEDPOTECQQCQRCR-----QESDP-----RQQYCQRRKEICE 38
Db 122 QQPSSAPQCQQCQNTCQQAAPYVQQCAPQCQQAFAACQQCOTSCQQTQCQQCCTPQQC 181

QY 39 E 39
Db 182 Q 182

RESULT 8
T29880 hypothetical protein F19G12.7 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29880
R;Nhan, M.; Le, T.T.
Submitted to the EMBL Data Library, April 1996
A;Description: The sequence of *C. elegans* cosmid F19G12.
A;Reference number: 220704
A;Accession: T29880
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-572 <NH>
A;Cross-references: EMBL:051997; PIDN:AAC48159.1; GSPDB:GN00028; CESP:F19G12.7
A;Experimental source: strain Bristol N2; clone F19G12
C;Genetics:
A;Gene: CESP:F19G12.7
A;Map position: X
A;Introns: 18/3; 160/3
C;Superfamily: ultra-high-sulfur keratin

Query Match 30.4%; Score 75.5; DB 2; Length 242;
Best Local Similarity 33.3%; Pred. No. 0.48; Mismatches 11; Indels 11; Gaps 4;
Matches 17; Conservative 8; Mismatches 7; Indels 18; Gaps 3;

QY 2 QEDPOTECQQCQRCR-----QESDP-RQQY--CORRKEICEEEE 42
Db 77 QQPAPARQCQQCQNTCQSAPVQQCAPCQQCAFACQ-QCQNNSCQQQQ 126

RESULT 9
T23681 hypothetical protein M02G9.1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23681
R;Matthews, L.

submitted to the EMBL Data Library, November 1996
A;Reference number: 219781
A;Accession: T23681
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1513 <WIL>
A;Cross-references: EMBL:Z81573; PIDN:CAR04625.1; GSPDB:GN00020; CESP:M02G9.1
A;Experimental source: clone M02G9
C;Genetics:
A;Gene: CESP:M02G9.1
A;Map position: 2
A;Introns: 23/3; 71/3; 121/3; 183/1; 108/3/3; 1141/3; 1408/1; 1449/1

Query Match 31.0%; Score 77; DB 2; Length 572;
Best Local Similarity 34.0%; Pred. No. 0.64; Mismatches 8; Indels 18; Gaps 3;
Matches 17; Conservative 8; Mismatches 7; Indels 18; Gaps 3;

QY 8 ECQCORRCRQ-----QESDPROQQCQRCKE--ICEEE 40
Db 381 QCQQCQNTCQOFAPVCEQQCASLCHQPSAPCQO-CQNTCQFAPVCEQQ 429

RESULT 10
T29699 hypothetical protein F31A3.1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T29699
R;Murray, J.; Le, T.T.
submitted to the EMBL Data Library, May 1996
A;Description: The sequence of *C. elegans* cosmid F31A3.
A;Reference number: 220667
A;Accession: T29699
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-242 <MR>
A;Cross-references: EMBL:058742; PIDN:AAB36856.1; GSPDB:GN00028; CESP:F31A3.1
A;Experimental source: strain Bristol N2; clone F31A3
C;Genetics:
A;Gene: CESP:F31A3.1
A;Map position: X
A;Introns: 18/3; 160/3
C;Superfamily: ultra-high-sulfur keratin

Query Match 30.4%; Score 75.5; DB 2; Length 242;
Best Local Similarity 33.3%; Pred. No. 0.48; Mismatches 11; Indels 11; Gaps 4;
Matches 17; Conservative 8; Mismatches 7; Indels 18; Gaps 3;

QY 2 QEDPOTECQQCQRCR-----QESDP-RQQY--CORRKEICEEEE 42
Db 77 QQPAPARQCQQCQNTCQSAPVQQCAPCQQCAFACQ-QCQNNSCQQQQ 126

RESULT 11
T31887 hypothetical protein C03A7.4 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T31887
R;Greco, T.; Bradshaw, H.; Elliott, G.
Submitted to the EMBL Data Library, July 1997
A;Description: The sequence of *C. elegans* cosmid C03A7.
A;Reference number: 221096
A;Accession: T31887
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-388 <GE>
A;Cross-references: EMBL:AF016451; PIDN:AAB66001.1; GSPDB:GN00023; CESP:C03A7.4
A;Experimental source: strain Bristol N2; clone C03A7
C;Genetics:
A;Gene: CESP:C03A7.4

A;Map position: 5
A;Introns: 75/3
C;Superfamily: gliadin

Query Match 30.2%; Score 75; DB 2; Length 388;
Best Local Similarity 37.5%; Pred. No. 0.77;
Matches 15; Conservative 7; Mismatches 16; Indels 2; Gaps 2;

Qy 2 QEDPQTEC-QQCQRRCRQEDPRQ-QQCQRREKEICEE 39
Db 194 QAQCOPQCPQQCQSCVQOOQNSQCEPANTQCSDICQQ 233

RESULT 14
T31898
hypothetical protein C03A7.7 - *Caenorhabditis elegans*
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T31898
R;Greco, T.; Bradshaw, H.; Elliott, G.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of *C. elegans* cosmid C03A7.
A;Reference number: Z21096
A;Accession: T31888
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Cross-references: EMBL:AF016451; PIDN:AAB66007.1; GSPPDB:GN00023; CESP:C03A7.7
A;Experimental source: strain Bristol N2; clone C03A7
C;Genetics:
A;Gene: CESP:C03A7.7
A;Map position: 5
A;Introns: 75/3
C;Superfamily: gliadin

Query Match 30.2%; Score 75; DB 2; Length 388;
Best Local Similarity 37.5%; Pred. No. 0.77;
Matches 15; Conservative 7; Mismatches 16; Indels 2; Gaps 2;

Qy 2 QEDPQTEC-QQCQRRCRQEDPRQ-QQCQRREKEICEE 39
Db 194 QAQCOPQCPQQCQSCVQOOQNSQCEPANTQCSDICQQ 233

RESULT 13
T31899
hypothetical protein C03A7.8 - *Caenorhabditis elegans*
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T31899
R;Greco, T.; Bradshaw, H.; Elliott, G.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of *C. elegans* cosmid C03A7.
A;Reference number: Z21096
A;Accession: T31899
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Cross-references: EMBL:AF016451; PIDN:AAB65995.1; GSPPDB:GN00023; CESP:C03A7.8
A;Experimental source: strain Bristol N2; clone C03A7
C;Genetics:
A;Gene: CESP:C03A7.8
A;Map position: 5
A;Introns: 75/3
C;Superfamily: gliadin

Query Match 29.8%; Score 74; DB 2; Length 438;
Best Local Similarity 30.2%; Pred. No. 1.1;
Matches 13; Conservative 12; Mismatches 12; Indels 6; Gaps 2;

Qy 3 EDPQTECQ-QCRRRC----RQQESDPQQYCQRRCKICEE 39
Db 191 QQAQAQCQPQQCQSCVQOOQNSQCEPANTQCSDICQQ 233

RESULT 14
T31898
hypothetical protein C03A7.14 - *Caenorhabditis elegans*
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T31898
R;Greco, T.; Bradshaw, H.; Elliott, G.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of *C. elegans* cosmid C03A7.
A;Reference number: Z21096
A;Accession: T31898
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-445 <GRE>
A;Cross-references: EMBL:AF016451; PIDN:AAB66007.1; GSPPDB:GN00023; CESP:C03A7.14
A;Experimental source: strain Bristol N2; clone C03A7
C;Genetics:
A;Gene: CESP:C03A7.14
A;Map position: 5
A;Introns: 18/3; 75/3
C;Superfamily: gliadin

Query Match 29.8%; Score 74; DB 2; Length 445;
Best Local Similarity 30.2%; Pred. No. 1.1;
Matches 13; Conservative 12; Mismatches 12; Indels 6; Gaps 2;

Qy 3 EDPQTECQ-QCRRRC----RQQESDPQQYCQRRCKICEE 39
Db 191 QQAQAQCQPQQCQSCVQOOQNSQCEPANTQCSDICQQ 233

RESULT 15
T44430
protein PV100 [imported] - winter squash
C;Species: Cucurbita maxima (winter squash)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T44430
R;Yamada, K.; Shimoda, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A;Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin
A;Reference number: Z22767; MUDB:99107919
A;Accession: T44430
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-810 <NM>
A;Cross-references: EMBL:AB019195; NID:93808061; PIDN:BAA34056.1; PID:93808062

Query Match 29.6%; Score 73.5; DB 2; Length 810;
Best Local Similarity 34.9%; Pred. No. 1.9;
Matches 15; Conservative 9; Mismatches 18; Indels 1; Gaps 1;

Qy 1 NQE-DPQTECQQCQRCRQEDPRQQYCQRRCKICEEE 42
Db 69 NQRGSPRAEYEVCRURCVAERGVFQQRKCEQVCEERLREREQ 111

Search completed: March 1, 2001, 15:52:37
Job Time: 561 sec